Rapid refinement of crystallographic protein construct definition employing enhanced hydrogen/deuterium exchange MS

Dennis Pantazatos*, Jack S. Kim*, Heath E. Klock†, Raymond C. Stevens‡, Ian A. Wilson‡, Scott A. Lesley†, and Virgil L. Woods, Jr.*§¶

*Department of Medicine, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA, 92093; ‡Joint Center for Structural Genomics, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037; and †Joint Center for Structural Genomics, Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive, San Diego, CA 92121

Communicated by S. Walter Englander‖, University of Pennsylvania School of Medicine, Philadelphia, PA, November 5, 2003 (received for review October 22, 2003)

Crystallographic efforts often fail to produce suitably diffracting protein crystals. Unstructured regions of proteins play an important role in this problem and considerable advantage can be gained in removing them. We have developed a number of enhancements to amide hydrogen/high-throughput and high-resolution deuterium exchange MS (DXMS) technology that allow rapid identification of unstructured regions in proteins. To demonstrate the utility of this approach for improving crystallization success, DXMS analysis was attempted on 24 *Thermotoga maritima* proteins with varying crystallization and diffraction characteristics. Data acquisition and analysis for 21 of these proteins was completed in 2 weeks and resulted in the localization and prediction of several unstructured regions within the proteins. When compared with those targets of known structure, the DXMS method correctly localized even small regions of disorder. DXMS analysis was then correlated with the propensity of such targets to crystallize and was further used to define truncations that improved crystallization. Truncations that were defined solely on DXMS analysis demonstrated greatly improved crystallization and have been used for structure determination. This approach represents a rapid and generalized method that can be applied to structural genomics or other targets in a high-throughput manner.

It is widely anticipated that access to high-resolution protein structures will be greatly facilitated by novel high-throughput improvements to conventional crystallographic methods. Proteome-scale crystallography is being pursued by several groups, including the Joint Center for Structural Genomics (JCSG) (1–3). These efforts have benefited greatly from recent technology enhancements in protein expression and crystallization. Despite these enhancements, successful production of stable proteins that can form x-ray analysis crystals continues to be a serious bottleneck. Many generally well structured proteins contain disordered regions that often serve as passive linkers between structurally autonomous domains, or become ordered when they interact with binding partners that provide stabilizing atomic contacts (4). Regardless of their function, unstructured regions can inhibit crystallization. Unstructured regions of proteins are also particularly susceptible to contaminating cellular proteases. Removing disordered regions may thus improve homogeneity. The energetics and kinetics of protein crystallization may be facilitated by selective deletion of these unstructured sequences (5). Even those proteins that readily crystallize can suffer from poor diffraction, and it is likely that disorder plays a significant role. Truncated constructs should result in better diffraction and, consequently, result in higher-resolution data more amenable to automated map-fitting procedures (6, 7).

In principle, information regarding protein dynamics could be used to design truncations that retain structure and maintain biological function but are otherwise depleted of disordered regions. A number of approaches ranging from stability-dependent protein expression screens to computation of stability from primary structure have been reported (8–10). For structural genomics studies, many targets have unknown folds, which limits the utility of bioinformatic predictions. NMR spectroscopy is one of the most powerful techniques to provide protein dynamics information; however, protein quantity, concentration, experimental time, and size are often limiting factors. Although limited proteolysis coupled to MS is a preferred approach, its use is time-consuming, frequently requiring that multiple proteolytic reactions be refined for optimal cleavage (6). Interpretation of limited proteolysis results is confounded by the possibility that proteolysis may clip internal loops, leading to destabilization and further proteolytic degradation of what originally was a structured region. Most importantly, there is no facile method to confirm that the truncations designed have retained the stable elements of the full-length protein. These approaches are problematic in structural genomics efforts, where high-throughput and low cost are dominating considerations (11).

For >40 years, peptide amide hydrogen-exchange techniques have been used to study the thermodynamics of protein conformational change and the mechanisms of protein folding (12, 13). More recently, they have proven to be increasingly powerful methods by which protein dynamics, domain structure, regional stability, and function can be studied (14, 15). Deuterium exchange methodologies coupled with liquid chromatography MS presently provide the most effective approach to study exchange rates in proteins (15). Proteolytic and/or collision-induced dissociation fragmentation methods allow exchange behavior to be mapped to subregions of the protein (15–26). Building on the pioneering work of Englander and coworkers (14, 15, 27), we have developed and implemented a number of improvements that have significantly improved throughput, comprehensiveness, and resolution. We term the method employing these enhancements high-throughput and high-resolution deuterium exchange MS (DXMS) (28–38).

Peptide amide hydrogens are not permanently attached to proteins, but they reversibly interchange with hydrogen present in solvent water. The chemical mechanisms of the exchange reactions are understood, and several well defined factors can profoundly alter exchange rates (12, 39–41). One of these factors is the extent to which a particular exchangeable hydrogen is exposed (accessible) to water. In a completely unstructured

Abbreviations: DXMS, high-throughput and high-resolution deuterium exchange MS; TM, *Thermatoga maritima* protein.

§To whom correspondence should be addressed at: Department of Medicine, University of California at San Diego, 9500 Gilman Drive, BSB 5078, La Jolla, CA 92093. E-mail: vwoods@ucsd.edu.

¶V.L.W. has an equity interest in ExSAR Corp.

‖S.W.E. is a consultant to ExSAR Corp.

BIOCHEMISTRY

Table 1. Description of *T. maritima* proteins studied, as classified by crystallization history

					Crystallization				Diffraction	
Target	Structure	SEG percent	DXMS percent	Location	Screened	Hits	Mountable	Percent/crystallized	Screened	Resolution, Å
TM0665	1J6N	11.2	3.1		1,920	247	244	25.6	7	2.2
TM1056	105J	17.7	0.0		3,360	175	153	9.8	45	1.8
TM0064	1J5S	3.0	1.1		2,880	114	146	9.0	31	1.8
TM1464	Pending	10.1	1.3		2,400	71	98	7.0	68	2.6
TM1080	1O1X	3.9	2.6		2,400	119	26	6.0	32	3.5
TM0449	1KQ4	2.6	6.4		1,152	42	16	5.0	16	2.3
TM0486		5.7	0.0		3,552	139	29	4.7	13	3.9
TM0542	Pending	9.3	2.9		2,592	47	49	3.7	56	2.8
TM1733		10.0	6.6	Internal	1,920	33	22	2.9	29	3.4
TM1158	1O1Y	2.5	2.6		2,880	32	39	2.5	62	2.2
TM0269	1J6R	2.8	4.0		4,608	55	33	1.9	11	2.1
TM1764		29.3	ND		480	4	2	1.3	0	
TM1816	1O13	11.0	17.7	Internal	2,016	2	21	1.1	9	2.0
TM0505		21.1	16.3	Internal	3,744	12	10	0.6	8	6.3
TM0212		21.3	0.0		1,152	5	1	0.5	1	8.6
TM0320		30.4	0.0		1,152	4	1	0.4	0	
TM1171		10.3	13.9	C terminus	2,400	4	1	0.2	0	
TM1171	D1	16.4	5.9		2,880	41	33	2.6	0	
TM1171	D2	16.5	5.9		2,880	66	9	2.6	16	2.3
TM1171	D3	15.3	5.6		1,920	24	10	1.8	0	
TM1171	D4 105L	14.9	5.6		1,920	12	14	1.4	1	2.1
TM1706		17.9	11.5	Internal	1,440	3	0	0.2	0	
TM1172		11.5	3.5	C terminus	2,208	3	1	0.2	0	
TM0913		7.5	2.2		4,320	5	2	0.2	0	
TM1079		16.3	5.7	Internal	1,920	3	0	0.2	0	
TM1773		10.4	ND		1,440	1	1	0.1	0	
TM0160		15.5	12.1	C terminus	2,400	2	1	0.1	0	
TM0160	D3 105Y	11.6	2.5		1,920	37	39	4.0	3	1.9
TM0855		9.1	ND		1,920	0	0	0.0	0	

Computational predictions employing the SEG algorithm to define segments of contrasting local compositional complexity (SEG percent) (48) and the portion of each protein's sequence found to be present in high exchange rate stretches of primary sequence (four or more rapidly exchanging contiguous residues; DXMS percent) are given as a percentage of total residues. The primary location of the DXMS-identified rapidly exchanging regions is indicated. The number of unique crystallization tests is indicated along with the number of tests showing crystal hits or crystals of sufficient size to mount for diffraction screening. The percentage of total tests that led to crystals is indicated. Those targets showing a $<1\%$ crystallization rate (all those in the lower section, except for the good crystallizing deletion constructs) are considered poorly crystallizing. The number of crystals screened for diffraction and the best resolution are indicated where data are available. ND, not determined.

polypeptide sequence, peptide amide hydrogens are always maximally accessible to water and exchange at their maximal rate, which is approximately (within a factor of 30) the same for all amides; their half-life of exchange is in the range of 1 sec at 0°C and pH 7.0 (41, 42). Most amide hydrogens in structured peptides or proteins exchange much more slowly (up to 10^9-fold reduction), reflecting the fact that exchange occurs only when transient unfolding fluctuations fully expose the amides to solvent water. The exception is the set of very-fast-exchanging amides in structured regions that have their amides fully solvent-exposed at all times, reflecting their protein surface disposition. In effect, each amide's exchange rate in a native protein directly and precisely reports solvent accessibility to it, thereby revealing the protein's thermodynamic stability on the scale of individual amino acids. Measurement of the exchange rates of a protein's amides can therefore allow direct identification and localization of structured/unstructured regions of the protein; unstructured regions are those where substantial contiguous stretches of primary sequence exhibit the maximal possible exchange rates, which is indicative of complete and continuous solvation of the amide hydrogens in such segments (12, 13). With its high-throughput capabilities, DXMS can rapidly localize disorder within crystallographic targets by using a minimum of protein sample.

The JCSG is a 5-year pilot project funded by the National Institutes of Health, National Institute of General Medical Sciences Protein Structure Initiative (PSI; www.nigms.nih.gov) to develop technologies for high-throughput protein structure determination. One aspect of these studies is focused on proteins from *Thermotoga maritima* (3). An unbiased set of *T. maritima* targets, 1,376 of the 1,877 predicted ORFs, were processed through expression and purification attempts. Of these ORFs, 542 proteins were expressed in soluble form and setup for crystallization trials with 434 resulting in preliminary crystal hits. This large data set provides the basis to select proteins for DXMS analysis based on their propensity to crystallize. To sharply focus this analysis, we have investigated a subset of *T. maritima* proteins selected for their range of known crystallization behavior. In the present study, we have used DXMS to improve crystallographic construct design under high-throughput conditions.

Methods

Protein Expression and Purification. Twenty-four *T. maritima* proteins were selected for analysis (see Table 1). These proteins, and the subsequently designed truncated constructs, were freshly prepared for this study as described (3). In brief, all targets were expressed in either *Escherichia coli* DL41 or HK100 from plasmids based on the expression vector pMH1 or pMH4. These

vectors encode a 12-aa tag containing the first six amino acids of thioredoxin and His_6 residues placed at the N terminus. Expression was induced by the addition of 0.15% arabinose for 3 h. Bacteria were lysed by sonication, cell debris were pelleted, and proteins were purified from the soluble fraction by nickel chelate chromatography. Proteins were concentrated to a final volume of 0.75 μl with concentrations ranging from 15 to 50 mg/ml in 20 mM Tris·HCl, pH 8.0/150 mM NaCl (3).

Establishment of Protein Fragmentation Probe Maps. Aliquots of each of the 24 proteins were adjusted to a concentration of 10 mg/ml in Tris-buffered saline [(TBS) 5 mM Tris-HCl/150 mM NaCl, pH 7.0], and all subsequent steps performed at 0°C on melting ice. In a 4°C cold room, 5 μl of each solution was further diluted with 15 μl of TBS in a microtiter plate employing multichannel pipettors for simultaneous manipulation. Thirty microliters of a stock exchange quench solution [0.8% formic acid, 1.6 M guanidine hydrochloride (Gdn·HCl) was then added to each sample (final concentration 0.5% formic acid, 1.0 M Gdn·HCl), samples were transferred to autosampler vials, and were then frozen on dry ice within 1 min after addition of quench solution as described (28–38). Vials with frozen samples were stored at −80°C until they were transferred to the dry ice-containing sample basin of the cryogenic autosampler module of a DXMS analysis apparatus designed and operated as described (36–38). In brief, samples were melted at 0°C, were proteolyzed for 16 sec by exposure to immobilized pepsin, and fragments were collected on a c18 HPLC column, with subsequent acetonitrile gradient elution. Column effluent was analyzed on both a Thermo Finnigan LCQ electrospray mass spectrometer and a Micromass Q-Tof mass spectrometer, as described (32–38). The SEQUEST software program (Thermo Finnigan, San Jose, CA) identified the likely sequence of the parent peptide ions and these tentative identifications were confirmed with specialized DXMS data reduction software as described (36–38).

On-Exchange Deuteration of Proteins. After establishment of fragmentation maps for each protein, amide hydrogen exchange-deuterated samples of each of the 24 proteins were prepared and processed exactly as above, except that 5 μl of each protein stock solution was diluted with 15 μl of deuterium oxide (D_2O) containing 5 mM Tris·HCl, 150 mM NaCl, pD (uncorrected reading) 7.0, and incubated for 10 sec at 0°C on melting ice before quenching and further processing. Data on the deuterated sample set were acquired in a single automated 30-h run and subsequent data reduction was performed with the DXMS software. Corrections for loss of deuterium label were made as described (36–38). The total time elapsed for data acquisition and analysis (both fragmentation maps and deuteration study) was 2 weeks. A total of 100 μg of each protein was used to complete the study. For subsequent comparative analysis of the exchange rates of amide hydrogens within truncated protein constructs versus their full-length forms, both proteins were contemporaneously on-exchanged as above, but were quenched at varying times (10, 30, 100, 300, 1,000, 3,000, 10,000, and 30,000 sec), and further processed as above, employing the fragmentation maps established for the full-length protein.

Protein Crystallization and Diffraction Data Acquisition. Proteins were crystallized by using the vapor diffusion method with 50 or 250 nl of protein and 50 or 250 nl of Mother liquor, with respective volumes as sitting drops on customized 96-well microtiter plates (Greiner, Nurtingen, Germany). Each protein was set up by using 480 commercial crystallization solutions [Wizard I/II and Wizard Cryo I/II (Emerald Biostructures, Bainbridge Island, WA), and Core Screen I/II, Cryo I, PEG ion, and Quad Grid (Hampton Research, Aliso Viejo, CA)] at 4°C and 20°C. Images of each crystal trial were taken at least twice, typically at 7 and 28 days after setup with an Optimag Veeco Oasis 1700 imager. Each image was evaluated by using a crystal detection algorithm and scored for the presence of crystals (44). Images at days 7 and 28 were also evaluated manually. Diffraction data were provided by the JCSG from automated data collection at 100 K on beamlines of the Stanford Synchrotron Radiation Laboratory Structural Molecular Biology/Macromolecular Crystallography Resource, and the Advanced Light Source beamlines 5.0.2 and 5.0.3 as described (3).

See *Supporting Methods*, which is published as supporting information on the PNAS web site, for further descriptions of the methods used in this study.

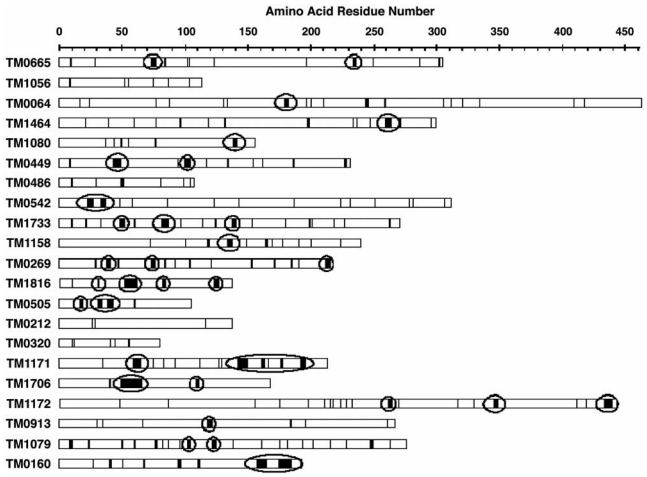

Fig. 1. The 10-sec deuteration results are shown for 21 proteins that were analyzed, whose amino acid lengths varied from 76 to 461 residues. Dark regions indicated fast-exchanging amides, and clear regions indicate stretches of no exchange. Regions of four or more fast-exchanging amides are circled. Corresponding boundaries for fast-exchanging amides are displayed in Fig. 5.

Results

DXMS Defines Rapidly Exchanging regions of *T. maritima* Proteins. In DXMS analysis, fragmentation parameters are initially optimized, including denaturant (Gdn·HCl) concentration, protease type(s), proteolysis duration to maximize the number of peptide fragment probes available for use with the target protein, and then the protein is examined by using a broad range of on-exchange times. This approach optimizes our ability to measure the widely ranging exchange rates for most of the peptide amides in the protein (32–38). In the present study, we sought to localize only disordered amides that exchanged very fast in the native protein. Based on prior experience, we used a single set of fragmentation conditions and on-exchanged samples for a single, brief (10 sec at 0°C) interval to selectively label only the most rapidly exchanging amides.

Generation of fragmentation maps and acquisition and analysis of deuteration data were completed in 2-weeks time for 24 samples. Fragmentation maps covering the entire protein sequence were obtained for 16 proteins, nearly complete coverage for five proteins, and inadequate coverage for three proteins (Table 1). Deuterium on-exchange studies were performed on the 21 proteins that had generated useful fragmentation maps (Fig. 1). Deuterium labeling was manually assigned to residue positions within the protein by first optimizing consensus in deuterium content of overlapping peptide probes, followed by further clustering of labeled amides together in the center of unresolved regions, so that a consensus map was generated. The deduced 10-sec exchange maps for each of the 21 proteins, and their consensus maps, are summarized in Fig. 1; the detailed deuterated fragment data are supplied in *Supporting Methods*.

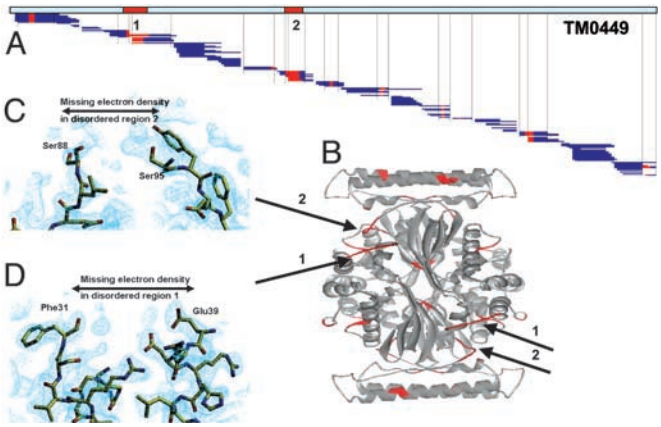

Fig. 2. (*A*) Ten-second amide hydrogen/deuterium exchange map for TM0449. The horizontal blue bars are the protein's pepsin-generated fragments that had been produced, identified, and used as exchange rate probes in the subsequent 10-sec deuteration study. The number of deuterons that went on to each peptide in 10 sec is indicated by the number of red residues in each peptide. Deuterium labeling was assigned to residue positions within the protein by first optimizing consensus in deuterium content of overlapping peptide probes, followed by further clustering of labeled amides together in the center of unresolved regions (with vertical bars indicating the range of possible location assignments), generating the consensus map (*Upper*), in which two extensive segments are seen to be deuterium-labeled: segment 1 (Phe-31-Glu-38) and segment 2 (Ser-88–Lys-93). (*B*) The electron density of the crystal indicates two regions of disordered sequence, corresponding to the segments 1 and 2. (*D* and *C*) Detailed electron density maps are shown, in which density is not visualized between the Phe-31–Glu-39 and Ser-88–Ser-95 regions of the TM0449 3D structure (45). DXMS-determined disorder constitutes 6.4% of this protein's sequence.

The duration of labeling (10 sec) was calculated to be sufficient to selectively deuterate primarily freely solvated amides (42, 43). This finding was confirmed by first fragmenting reference proteins with pepsin to yield unstructured peptides, followed by deuterium-exchange labeling of the resulting peptide mix for 10 sec at 0°C, pH 7.0, as above, and then quenching and subjecting the mixture to DXMS analysis, but without repeated proteolysis. Under these conditions, all peptides were saturation-labeled with a 10-sec period of on-exchange (data not shown).

DXMS Correctly Localized Disordered Regions in Control Proteins with Known 3D Structures. Interpretation of the exchange maps of the *T. maritima* proteins was guided by the expectation of two patterns of fast-exchange labeling: structurally stable, but well solvated, rapidly exchanging residues (one to three contiguous residues) versus labeling of longer stretches of sequence (four or more residues) indicative of disorder. It was presumed that three contiguous amino acids was likely the smallest number needed to complete a structurally stable turn on the surface of a protein. The percent of each protein's residues that rapidly labeled in stretches of four or more residues is indicated as DXMS percent in Table 1.

The structure of *T. maritima* thy1 protein TM0449 has been determined to 2.25 Å (45). Its exchange map demonstrated two segments (more than or equal to four residues in each) with rapid exchange, labeled A (Phe-31–Glu-38) and B (Ser-88–Lys-93), and several isolated rapidly exchanging amides in groups of three or less, scattered throughout the sequence (Fig. 2*A*). Both of the rapidly exchanging segments corresponded closely to regions of disorder in the crystal (Phe-32–Glu-38 and Ser 89-Ser 94; Fig. 2*B*) confirming the ability of DXMS data to detect and localize such disordered regions. Interestingly, these regions also appear to be involved in the binding of the enzyme substrate and adopt a structured conformation after binding ligand (45). This finding

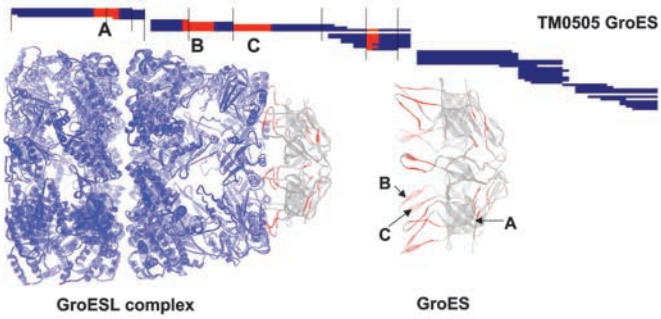

Fig. 3. The on-exchange map of TM0505 indicates three internal segments (A, B, and C) of rapidly exchanging amides. The internal segments are mapped onto the crystal structure of the GroES protein homolog of TM0505. The *M. tuberculosis* GroEL subunit is blue, and the heptamer complex of *M. tuberculosis* GroES subunits is gray. The homologous locations of rapid exchange sites in the *T. maritima* protein are red. Disorder constitutes 16.3% of this protein's sequence.

suggests that DXMS can also provide some localized prediction of substrate and cofactor binding sites. This result raises the caution that even focused deletion of unstructured regions always carries the potential to remove regions critical to biological function. Similar comparisons were performed for other proteins with known structures (data not shown) with regions of internal disorder typically mapping to loop or extended solvent-accessible regions.

Poorly Crystallizing *T. maritima* Proteins Contain Substantial Disorder. The exchange map for *T. maritima* GroES heat shock protein TM0505 demonstrated rapid exchange for three segments containing four or more contiguous rapidly exchanging residues, which together constitute 16% of its sequence (Fig. 3). Whereas this *T. maritima* protein had previously produced only poorly diffracting crystals, it is a close homolog of the GroES heat shock protein of *Mycobacterium tuberculosis*, for which crystal structures were available as the GroES heptamer, and as a complex (GroELS) with the GroEL subunit (46, 47). When the *T. maritima* residues with rapid exchange are mapped on the *M. tuberculosis* structures, they predominantly localize to disordered residues in GroES that make contact with the GroEL-binding surface.

The exchange map for the conserved hypothetical protein TM1816 (Fig. 1, and Fig. 5, which is published as supporting information on the PNAS web site) is dominated by several substantial regions of disorder, constituting 17.7% of its residues. This protein was a unique example where a structure was obtained from a target exhibiting substantial disorder. The poorly crystallizing proteins TM1171, TM0160, TM1706, TM1733, and TM1079 exhibit, for substantial portions of their sequence, rapidly exchanging stretches of four or more residues (13.9%, 12.1%, 11.5%, 6.6%, and 5.7%, respectively; Fig. 1 and *Supporting Methods*). TM0160, TM1171, and TM1172 had disorder primarily at the C terminus; (Fig. 1 and *Supporting Methods*). These targets offer a straightforward route to domain optimization by simple deletion of the disordered C terminus. The optimization of two of these targets is described below.

Disorder-Depleted Constructs of *T. maritima* Proteins Preserve Ordered Structure. Truncation mutants of TM0160 and TM1171 proteins were prepared (Fig. 4 *A* and *B*), in which the C-terminal disordered region(s) of both proteins were deleted. The fragmentation patterns produced by pepsin often exhibited preferences for sites near exchange-defined stretches of disorder. We therefore produced several truncated constructs to each full-length protein, in part guided by the location of the "preferred"

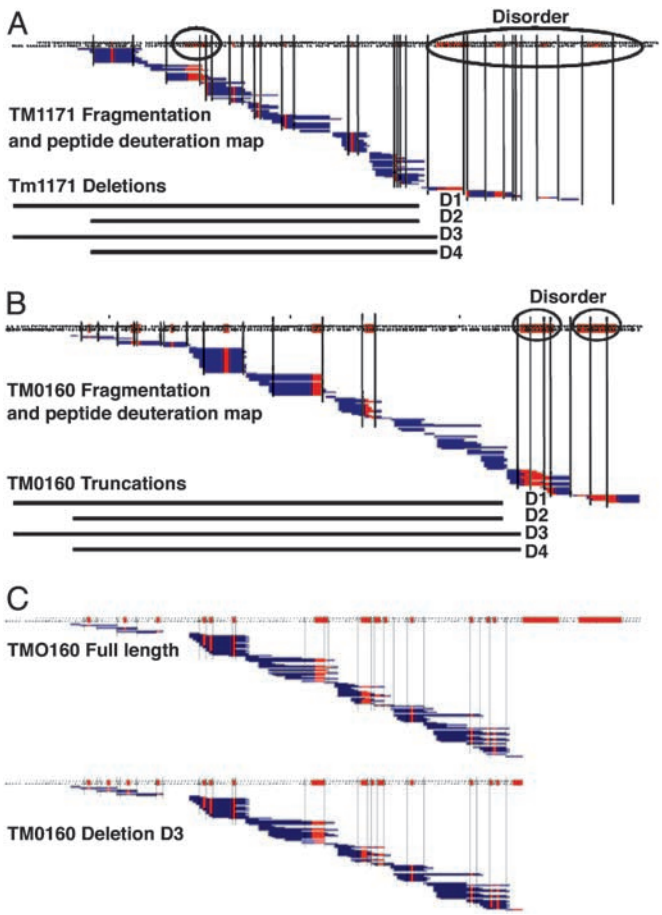

Fig. 4. TM1171 (*A*) and TM0160 (*B*) show substantial C-terminal disorder (circled sequences). Four truncated constructs of each protein were made by eliminating the C-terminal regions (D1–D4). (*C*) Repeat DXMS analysis demonstrates that deletion constructs of TM0160 preserve the core full-length structure. Full-length TM0160, and its longest truncation (D3), were on-exchanged variously for 10, 100, 1,000, and 10,000 sec at 0°C, were exchange-quenched, and were subjected to comparative DXMS analysis as described above. The resulting comprehensive exchange maps for full-length (*Upper*) and D3 truncated (*Lower*) had virtually identical patterns (10-sec exchange time is shown).

pepsin cut sites, and for both TM0160 and TM1171. Deletions were designed solely on the basis of DXMS experimental data. The truncations expressed well as a soluble protein. Full-length TM0160, and its longest truncated version (D3), were on-exchanged variously for 10, 100, 1,000, and 10,000 sec at 0°C on ice, were exchange-quenched, and subjected to comparative DXMS analysis as described above. The resulting 10-sec exchange maps for full-length protein and the D3 truncated version (Fig. 4*C*) had virtually identical 10-sec patterns, and detailed analysis of the longer exchange times demonstrated that D3 had a stability profile identical to that of the TM0160 full-length. Similarly, each of the four TM1171-truncated constructs expressed well as soluble protein, and had DXMS stability maps identical to that of the TM1171 full-length protein in the corresponding sequence regions (data not shown).

Deletion Constructs of Two *T. maritima* Proteins Show Marked Improvement in Crystallization. Full-length TM0160 and the D3 truncation were submitted for crystallization trials (Table 1). A total of 480 commercially available crystallization solutions were screened at 4°C and 20°C as described in *Methods*. From multiple protein preparations and crystallization attempts, the full-length protein showed marginal crystals (inadequate for diffraction experiments) for only 3 of 2,400 total attempts. In contrast, by using the same 480 crystallization solutions, 76 crystal hits were obtained for the truncated constructs from 1,920 attempts. Crystals from the TM0160 D3 truncation mutant had better morphology than the few crystals obtained with the full-length construct, and diffracted well. Ultimately, a 1.9-Å data set from selenomethionine-incorporated protein enabled determination of the TM0160 3D structure, which represents another fold (to be presented elsewhere). Similarly, the TM1171 and truncations were subjected to crystallization trials. Whereas the TM1171 full-length protein again showed very marginal crystallization propensity (5 of 2,400 attempts), each of the four TM1171 deletion constructs showed marked improvement in crystallization success with the TM1171-D4 construct ultimately resulting in a 2.1-Å data set that was used to determine its 3D structure (to be presented elsewhere). It should be noted that good diffracting crystals were obtained for DXMS-designed deletion constructs in both native and selenomethionine forms.

Discussion

These studies have shown that DXMS analysis can reliably detect and localize disordered regions within an otherwise structured protein. Stability profiles were determined for 21 *T. maritima* proteins that had previously been subjected to crystallization studies (Table 1). Twelve proteins crystallized readily in >1% of the conditions tested. Four of the remaining nine poorly crystallizing proteins had a high fraction (>10%) of their sequence in disordered regions, suggesting a potential cause of the poor behavior. Most importantly, our present instrumentation allowed determination of the DXMS protein-stability profiles at speeds matching the needs of high-throughput structural genomics.

We have also established that successful strategies to selectively delete disorder from protein constructs can be readily discerned from DXMS stability profiles. Furthermore, we have shown that DXMS can rapidly and reliably assess the preservation of the structure of the ordered regions/domains in the full-length wild-type protein in the resulting truncations. Whereas several bioinformatic approaches to construct design can be used with well characterized protein folds, DXMS- guided construct redesign offers a particular advantage in the study of proteins that have novel folds. DXMS data directly localize disorder to specific amino acid residues in the target protein regardless of overall fold structure, allowing greatly refined truncation definition. Unlike NMR methods, which can also provide exchange data, DXMS requires only microgram amounts of soluble protein and data acquisition and analysis can be performed in a rapid time scale. In the present investigation, the total time elapsed for data acquisition and analysis (both fragmentation maps and deuteration study) was 2 weeks, and a total of 100 μg of each protein was used.

Finally, these results establish that DXMS stability profile-guided construct design can produce derivatives of poorly crystallizing proteins that crystallize and diffract well. In each of two attempts, we have succeeded in producing diffraction quality crystals of truncated constructs of *T. maritima* full-length proteins that had behaved poorly in several prior crystallization trials, and have confirmed that these truncations preserved full-length exchange rate patterns, indicating that they had retained wild-type structure with high fidelity. Taken together, these results indicate that DXMS is a valuable tool for structural genomics efforts.

We thank S. Walter Englander, David Wemmer, and Philip Bourne for their support and guidance in this DXMS application; Linda Brinen, Glen Spraggon, and Ashley Deacon for providing diffraction screening

data; Dan McMullan, Kevin Rodrigues, Juli Vincent, and Eileen Ambing for protein purification and crystallization studies; and Andreas Kreusch for assistance with figures. This work was supported by National Institutes of Health Grant CA099835 (to V.L.W.) and by National Institutes of Health Protein Structure Initiative Grant GM-99-009, National Institute of General Medical Sciences Grant GM 062411 (to I.A.W.), and grants from the University of California BioStar and Life Sciences Informatics Programs Grants S97-90, S99-44, and L98-30 (to V.L.W.), with the matching corporate sponsor for these grants being ExSAR Corporation (Monmouth Junction, NJ).

1. Stevens, R. C. & Wilson, I. A. (2001) *Science* **293,** 519–520.
2. Stevens, R. C., Yokoyama, S. & Wilson, I. A. (2001) *Science* **294,** 89–92.
3. Lesley, S. A., Kuhn, P., Godzik, A., Deacon, A. M., Mathews, I., Kreusch, A., Spraggon, G., Klock, H. E., McMullan, D., Shin, T., *et al.* (2002) *Proc. Natl. Acad. Sci. USA* **99,** 11664–11669.
4. Wright, P. E. & Dyson, H. J. (1999) *J. Mol. Biol.* **293**, 321–331.
5. Kwong, P., Wyatt, R., Desjardins, E., Robinson, J., Culp, J., Hellmig, B., Sweet, R., Sodroski, J. & Hendrickson, W. (1999) *J. Biol. Chem.* **274,** 4115–4123.
6. Cohen, S. L., Ferre-D'Amare, A. R., Burley, S. K. & Chait, B. T. (1995) *Protein Sci.* **4,** 1088–1099.
7. Lamzin, V. S. & Perrakis, A. (2000) *Nat. Struct. Biol.* **7,** Suppl., 978–981.
8. Dunker, A. K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., Obradovic, Z., Kissinger, C. & Villafranca, J. E. (1998) *Pac. Symp. Biocomput.* **3,** 473–484.
9. Garner, E., Cannon, P., Romero, P., Obradovic, Z. & Dunker, A. K. (1998) *Genome Inform.* **9,** 201–214.
10. Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., Garner, E., Guilliot, S. & Dunker, A. K. (1998) *Pac. Symp. Biocomput.* **3,** 437–448.
11. Chen, G.-Q., Sun, Y., Jin, R. & Gouaux, E. (1998) *Protein Sci.* **7,** 2623–2630.
12. Englander, S. W. & Englander, J. J. (1994) *Methods Enzymol.* **232,** 26–42.
13. Bai, Y., Englander, J. J., Mayne, L., Milne, J. S. & Englander, S. W. (1995) *Methods Enzymol.* **259,** 344–356.
14. Englander, S. W., Mayne, L., Bai, Y. & Sosnick, T. R. (1997) *Protein Sci.* **6,** 1101–1109.
15. Engen, J. R. & Smith, D. L. (2001) *Anal. Chem.* **73,** 256A–265A.
16. Hoofnagle, A. N., Resing, K. A., Goldsmith, E. J. & Ahn, N. G. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 956–961.
17. Resing, K. A., Hoofnagle, A. N. & Ahn, N. G. (1999) *J. Am. Soc. Mass Spectrom.* **10,** 685–702.
18. Mandell, J. G., Falick, A. M. & Komives, E. A. (1998) *Anal. Chem.* **70,** 3987–3995.
19. Mandell, J. G., Falick, A. M. & Komives, E. A. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 14705–14710.
20. Mandell, J. G., Baerga-Ortiz, A., Akashi, S., Takio, K. & Komives, E. A. (2001) *J. Mol. Biol.* **306,** 575–589.
21. Kim, M. Y., Maier, C. S., Reed, D. J. & Deinzer, M. L. (2001) *J. Am. Chem. Soc.* **123,** 9860–9866.
22. Kim, M. Y., Maier, C. S., Reed, D. J., Ho, P. S. & Deinzer, M. L. (2001) *Biochemistry* **40,** 14413–14421.
23. Zhang, Y. H., Yan, X., Maier, C. S., Schimerlik, M. I. & Deinzer, M. L. (2001) *Protein Sci.* **10,** 2336–2345.
24. Kim, M. Y., Maier, C. S., Reed, D. J. & Deinzer, M. L. (2002) *Protein Sci.* **11,** 1320–1329.
25. Peterson, V. J., Barofsky, E., Deinzer, M. L., Dawson, M. I., Feng, K. C., Zhang, X. K., Madduru, M. R. & Leid, M. (2002) *Biochem. J.* **362,** 173–181.
26. Yan, X., Zhang, H., Watson, J., Schimerlik, M. I. & Deinzer, M. L. (2002) *Protein Sci.* **11,** 2113–2124.
27. Smith, D. L., Deng, Y. & Zhang, Z. (1997) *J. Mass Spectrom.* **32,** 135–146.
28. Woods, V. L., Jr. (1997) U.S. Patent 5,658,739.
29. Woods, V. L., Jr. (2001) U.S. Patent 6,291,189.
30. Woods, V. L., Jr. (2001) U.S. Patent 6,331,400.
31. Woods, V. L., Jr. (2003) U.S. Patent 6,599,707.
32. Hamuro, Y., Burns, L. L., Canaves, J. M., Hoffman, R. C., Taylor, S. S. & Woods, V. L., Jr. (2002) *J. Mol. Biol.* **4,** 703–714.
33. Hamuro, Y., Wong, L., Shaffer, J., Kim, J. S., Jennings, P. A., Adams, J. A. & Woods, V. L., Jr. (2002) *J. Mol. Biol.* **323,** 871–881.
34. Woods, V. L., Jr., & Hamuro, Y. (2001) *J. Cell. Biochem.* **37,** 89–98.
35. Woods, V. L., Jr. (2003) U.S. Patent 6,599,707.
36. Zawadzki, K. M., Hamuro, Y., Kim, J. S., Garrod, S., Stranz, D. D., Taylor, S. S. & Woods, V. L., Jr. (2003) *Protein Sci.* **12,** 1980–1990.
37. Englander, J. J., DelMar, C., Li, W., Englander, S. W., Kim, J. S., Stranz, D. D., Hamuro, Y. & Woods, V. L., Jr. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 7057–7062.
38. Hamuro, Y., Zawadzki, K. M., Kim, J. S., Stranz, D., Taylor, S. S. & Woods, V. L., Jr. (2003) *J. Mol. Biol.* **327,** 1065–1076.
39. Englander, J. J., Rogero, J. R. & Englander, S. W. (1985) *Anal. Biochem.* **147,** 234–244.
40. Englander, S. W. & Englander, J. J. (1972) *Methods Enzymol.* **26,** 406–413.
41. Englander, S. W. & Englander, J. J. (1978) *Methods Enzymol.* **49,** 24–39.
42. Molday, R. S., Englander, S. W. & Kallen, R. G. (1972) *Biochemistry* **11,** 150–158.
43. Bai, Y., Milne, J. S., Mayne, L. C. & Englander, S. W. (1993) *Proteins Struct. Funct. Genet.* **17,** 74–86.
44. Spraggon, G., Lesley, S. A., Kreusch, A. & Priestle, J. P. (2002) *Acta Crystallogr. D* **58,** 1915–1923.
45. Mathews, I. I., Deacon, A. M., Canaves, J. M., McMullan, D., Lesley, S. A., Agarwalla, S. & Kuhn, P. (2003) *Structure (London)* **11,** 677–690.
46. Ranson, N. A., Farr, G. W., Roseman, A. M., Gowen, B., Fenton, W. A., Horwich, A. L. & Saibil, H. R. (2001) *Cell* **107,** 869–879.
47. Roberts, M. M., Coker, A. R., Fossati, G., Mascagni, P., Coates, A. R. M. & Wood, S. P. (2003) *J. Bacteriol.* **185,** 4172–4185.
48. Wootton, J. (1994) *Comput. Chem.* **18,** 269–285.